

CV - Yutong Shao

1. Personal Information

Personal Webpage: sythello.github.io

2. Education

09/2018 – 12/2023 Computer Science and Engineering Department, UC San Diego

Degree: Ph.D.

Major: Computer Science (GPA: 4.0/4.0)

Advisor: Prof. Ndapa Nakashole

09/2014 - 07/2018 School of Electronics Engineering and Computer Science,
Peking University

Degree: B.S., with honor

Major: Computer Science (GPA: 3.73/4.0; ranking: 5/190)

Advisor: Prof. Zhi-Hong Deng

3. Work History

2024-(now)	Machine Learning Engineer, Tiktok	San Jose, California
2022	Applied Scientist Intern, Amazon Alexa	Seattle, Washington
2021	Research Scientist Intern, Megagon Labs	Mountain View, California
2017	Research Intern, University of Edinburgh	Edinburgh, Scotland, UK

4. Publications

- (1) **Y. Shao**, A. Kumar, N. Nakashole. *Database-aware ASR Error Correction for Speech-to-SQL Parsing*, ICASSP 2023.
- (2) **Y. Shao**, A. Kumar, and N. Nakashole. *Structured Data Representation in Natural Language Interfaces*, IEEE Data Eng. Bull. 2022 (invited paper).
- (3) **Y. Shao**, N. Bhutani, S. Rahman, E. Hruschka. *Low-resource Entity Set Expansion on User-generated Text*, NAACL 2022 (Findings). (Work done during internship at Megagon Labs)
- (4) Y. Wang, **Y. Shao**, N. Nakashole. *Interactive Plot Manipulation using Natural Language*, NAACL 2021 (Demo).
- (5) **Y. Shao**, N. Nakashole. *ChartDialogs: Plotting from Natural Language Instructions*, ACL 2020.
- (6) **Y. Shao**, R. Sennrich, B. L. Webber, and F. Fancellu. *Evaluating Machine*

Translation Performance on Chinese Idioms with a Blacklist Method, LREC 2018.
(Work done during internship at University of Edinburgh)

5. Other Academic Experience

- (1) A Study on Linearizing Structured Data: Insights from Text-to-SQL, 2023. Paper under submission at ARR.
 - *We conducted detailed explorations into the inner mechanisms of a strong text-to-SQL model with T5 backbone on linearized structured data, obtaining various interesting findings on its internal behaviors.*
- (2) Multi-document Product Review Summarization, applied scientist internship at Amazon Alexa AI, 2022.
 - *Improved LongFormer-based summarization model using sentence clustering and distant-supervised fine-tuning of BERT-based aspect extractor.*
- (3) Exploring Domain Adaptability for Sentiment Classification Models, open-ended course project, 2018.
 - *A comparative study to evaluate the domain transfer abilities of various representative sentiment classification models, neural and non-neural.*
- (4) Stock Price Prediction Based on Deep Learning, Bachelor's thesis supervised by Prof. Zhi-Hong Deng, Spring 2018.
 - *Developed a working method for stock prediction using LSTM-VAE on sequential features extracted from raw prices, human-designed metrics and textual news.*
- (5) Automatic Headline Generation based on Deep Learning, undergraduate research program under the supervision of Prof. Zhi-Hong Deng, Autumn 2017.
 - *Developed a Chinese news headline generation model based on hierarchical attention network (HAN) and LSTM.*
- (6) SemEval Contest, Task 1 (Semantic Textual Similarity), Autumn 2016, ranked 8/32 in the En-En subtask.
 - *Developed a sentence similarity prediction model based on token-wise embedding similarity matrix and CNN.*

6. Honors

- | | |
|------|---|
| 2017 | Merit Student, Peking University |
| 2017 | Outstanding Undergraduate Research Project, Peking University |
| 2017 | The Third Prize, ACM Campus Contest, Peking University |
| 2016 | TP-Link Scholarship, Peking University |
| 2016 | The Third Prize, ACM Campus Contest, Peking University |
| 2015 | Suzhou Industrial Park Scholarship, Peking University |
| 2015 | The Second Prize, ACM Campus Contest, Peking University |

2015 The Second Prize, "Jiangzehan" Mathematical Modeling Contest, Peking University

7. Technical Skills

Proficient Programming Language: Python, Java, C++

Proficient Software & Packages: PyTorch, AllenNLP, Matplotlib, Amazon MTurk

8. Teaching Experience

- TA for course "Statistical Natural Language Processing" (CSE 256) at UCSD, Spring 2023.
- TA for course "Recommender Systems and Web Mining" (CSE 158) at UCSD, Winter 2023.
- TA for course "Topics in Computer Science & Engineering: Advanced Statistical NLP (Research)" (CSE 291) at UCSD, Winter 2021.
- TA for course "Algorithm Design and Analysis" (Discussion Sessions) at Peking University, 2017.